

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.**

Application Serial Number: 10/540,310  
Source: PCT  
Date Processed by STIC: 06-30-05

# ***ENTERED***



PCT

## RAW SEQUENCE LISTING

DATE: 06/30/2005

PATENT APPLICATION: US/10/540,310

TIME: 08:45:07

Input Set : A:\689290-244.txt

Output Set: N:\CRF4\06302005\J540310.raw

```

3 <110> APPLICANT: Avalon Pharmaceuticals, Inc.
5 <120> TITLE OF INVENTION: Breast Specific Protein Expressed in Cancer and Methods of
6   Use Thereof
8 <130> FILE REFERENCE: 689290-183
C--> 10 <140> CURRENT APPLICATION NUMBER: US/10/540,310
C--> 11 <141> CURRENT FILING DATE: 2005-06-20
13 <150> PRIOR APPLICATION NUMBER: US/60/434,960
14 <151> PRIOR FILING DATE: 2002-12-20
16 <160> NUMBER OF SEQ ID NOS: 4
18 <170> SOFTWARE: PatentIn version 3.0
20 <210> SEQ ID NO: 1
21 <211> LENGTH: 629
22 <212> TYPE: DNA
23 <213> ORGANISM: Homo sapiens
25 <220> FEATURE:
26 <221> NAME/KEY: misc_feature
27 <222> LOCATION: (1)..(629)
28 <223> OTHER INFORMATION: n=a, t, g, or c
31 <400> SEQUENCE: 1
32 tgtggcatat tatttttatt atgggggaaag ataactgaga ataaagctat catgcagata      60
34 tttgcagaga taaaagtaat gcagatactg agtggagttt tgatcaaact atgcttgaaa      120
36 gccactctac cactagttac acaaaccaat aatttcctt cgcagtggaa gtcagcttga      180
38 gttttttcag gtgtttttgt gggtttcacc agatacagca aagaaattaa aattactgtt      240
40 aatggatgtc aaaaccagtc agaagtatcc taagttatat aatttgtcaa acaaccatat      300
42 acatatattt tgtattatat ttatcctttt gttcttcctt tggtaggaaa attgtctcat      360
44 taattcttat acgaaaggac ttaaaattag caaacttttt ttgcaaacac atggattcca      420
46 ttcttgagact tgaggacaac ttgacgaaca ggctggggag gccttgagtg gtctggagcc      480
W--> 48 agcttggaagc ggagcagagt taatgccact gccactntac actcaattat ggcaaaatgc      540
50 tgcccaatgc agttccttaa tccagctgag aatggtatga aggcatangg atgtatnntt      600
52 tcagaaattt ccctggagaa tctcagggg                                629
55 <210> SEQ ID NO: 2
56 <211> LENGTH: 760
57 <212> TYPE: PRT
58 <213> ORGANISM: Homo sapiens
60 <400> SEQUENCE: 2
62 Gly Gly Ala Ala Ala Gly Ala Thr Ala Ala Cys Thr Gly Ala Gly Ala
63 1           5           10           15
65 Ala Thr Ala Ala Ala Gly Cys Thr Ala Thr Cys Ala Thr Gly Cys Ala
66           20           25           30
68 Gly Ala Thr Ala Thr Thr Thr Gly Cys Ala Gly Ala Gly Ala Thr Ala
69           35           40           45
71 Ala Ala Ala Gly Thr Ala Ala Thr Gly Cys Ala Gly Ala Thr Ala Cys
72           50           55           60

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/540,310

DATE: 06/30/2005

TIME: 08:45:07

Input Set : A:\689290-244.txt

Output Set: N:\CRF4\06302005\J540310.raw

```

74 Thr Gly Ala Gly Thr Gly Gly Ala Gly Thr Thr Thr Thr Gly Ala Thr
75 65                                70                                75                                80
77 Cys Ala Ala Ala Cys Thr Ala Thr Gly Cys Thr Thr Gly Ala Ala Ala
78                                85                                90                                95
80 Gly Cys Cys Ala Cys Thr Cys Thr Ala Cys Cys Ala Cys Thr Ala Gly
81                                100                               105                               110
83 Thr Thr Ala Cys Ala Cys Ala Ala Ala Cys Cys Ala Ala Thr Ala Ala
84                                115                               120                               125
86 Thr Thr Thr Cys Cys Cys Thr Thr Cys Gly Cys Ala Gly Thr Gly Gly
87                                130                               135                               140
89 Ala Ala Gly Thr Cys Ala Gly Cys Thr Thr Gly Ala Gly Thr Thr Thr
90 145                               150                               155                               160
92 Thr Thr Thr Cys Ala Gly Gly Thr Gly Thr Thr Thr Thr Thr Gly Thr
93                                165                               170                               175
95 Gly Gly Gly Thr Thr Thr Cys Ala Cys Cys Ala Gly Ala Thr Ala Cys
96                                180                               185                               190
98 Ala Gly Cys Ala Ala Ala Gly Ala Ala Ala Thr Thr Ala Ala Ala Ala
99                                195                               200                               205
101 Thr Thr Ala Cys Thr Gly Thr Thr Ala Ala Thr Gly Gly Ala Thr Gly
102                                210                               215                               220
104 Thr Cys Ala Ala Ala Ala Cys Cys Ala Gly Thr Cys Ala Gly Ala Ala
105 225                               230                               235                               240
107 Gly Thr Ala Thr Cys Cys Thr Ala Ala Gly Thr Thr Ala Thr Ala Thr
108                                245                               250                               255
110 Ala Ala Thr Thr Thr Gly Thr Cys Ala Ala Ala Cys Ala Ala Cys Cys
111                                260                               265                               270
113 Ala Thr Ala Thr Ala Cys Ala Thr Ala Thr Ala Thr Thr Thr Thr Gly
114                                275                               280                               285
116 Thr Ala Thr Thr Ala Thr Ala Thr Thr Thr Ala Thr Cys Cys Thr Thr
117                                290                               295                               300
119 Thr Thr Gly Thr Thr Cys Thr Thr Cys Cys Thr Thr Thr Gly Gly Thr
120 305                               310                               315                               320
122 Ala Gly Gly Ala Ala Ala Ala Thr Thr Gly Thr Cys Thr Cys Ala Thr
123                                325                               330                               335
125 Thr Ala Ala Thr Thr Cys Thr Thr Ala Thr Ala Cys Gly Ala Ala Ala
126                                340                               345                               350
128 Gly Gly Ala Cys Thr Thr Ala Ala Ala Ala Thr Thr Ala Gly Cys Ala
129                                355                               360                               365
131 Ala Ala Cys Thr Thr Thr Thr Thr Thr Thr Gly Cys Ala Ala Ala Cys
132                                370                               375                               380
134 Ala Cys Ala Thr Gly Gly Ala Thr Thr Cys Cys Ala Thr Thr Cys Thr
135 385                               390                               395                               400
137 Thr Gly Gly Ala Cys Thr Thr Gly Ala Gly Gly Ala Cys Ala Ala Cys
138                                405                               410                               415
140 Thr Thr Gly Ala Cys Gly Ala Ala Cys Ala Gly Gly Cys Thr Gly Gly
141                                420                               425                               430
143 Gly Gly Ala Gly Gly Cys Cys Thr Thr Gly Ala Gly Thr Gly Gly Thr
144                                435                               440                               445
146 Cys Thr Gly Gly Ala Gly Cys Cys Ala Gly Cys Thr Thr Gly Ala Ala

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/540,310

DATE: 06/30/2005

TIME: 08:45:07

Input Set : A:\689290-244.txt

Output Set: N:\CRF4\06302005\J540310.raw

```

147      450      455      460
149 Gly Cys Gly Gly Ala Gly Cys Ala Gly Ala Gly Thr Thr Ala Ala Thr
150 465      470      475      480
152 Gly Cys Cys Ala Cys Thr Gly Cys Cys Ala Cys Thr Thr Thr Ala Cys
153      485      490      495
155 Ala Cys Thr Cys Ala Ala Thr Thr Ala Thr Gly Gly Cys Ala Ala Ala
156      500      505      510
158 Ala Thr Gly Cys Thr Gly Cys Cys Cys Ala Ala Thr Gly Cys Ala Gly
159      515      520      525
161 Thr Thr Cys Cys Thr Thr Ala Ala Thr Cys Cys Ala Gly Cys Thr Gly
162      530      535      540
164 Ala Gly Ala Ala Thr Gly Thr Ala Thr Gly Ala Ala Gly Gly Cys
165 545      550      555      560
167 Ala Thr Ala Ala Gly Ala Thr Gly Thr Ala Thr Thr Thr Thr Thr
168      565      570      575
170 Cys Ala Gly Ala Thr Thr Thr Thr Cys Cys Cys Thr Thr Gly Ala Gly
171      580      585      590
173 Ala Ala Thr Cys Thr Cys Ala Ala Gly Cys Gly Thr Thr Ala Ala Ala
174      595      600      605
176 Gly Ala Cys Cys Thr Gly Ala Gly Gly Gly Gly Thr Cys Thr Cys Cys
177      610      615      620
179 Cys Ala Gly Ala Ala Ala Thr Ala Gly Gly Gly Gly Thr Thr Thr Gly
180 625      630      635      640
182 Gly Gly Thr Gly Ala Ala Gly Ala Gly Cys Cys Cys Ala Cys Ala Thr
183      645      650      655
185 Ala Thr Thr Gly Ala Thr Ala Ala Cys Ala Cys Ala Gly Thr Ala Thr
186      660      665      670
188 Thr Cys Cys Thr Gly Cys Cys Cys Gly Thr Ala Cys Gly Ala Gly Cys
189      675      680      685
191 Gly Thr Cys Cys Ala Thr Cys Thr Gly Gly Ala Ala Ala Gly Gly
192      690      695      700
194 Ala Ala Thr Gly Gly Gly Thr Thr Thr Gly Cys Gly Ala Gly Thr Ala
195 705      710      715      720
197 Cys Ala Cys Gly Gly Ala Thr Ala Thr Thr Gly Thr Thr Ala Cys Thr
198      725      730      735
200 Thr Ala Cys Cys Gly Thr Gly Cys Gly Thr Ala Asn Ala Ala Gly Cys
201      740      745      750
203 Gly Gly Ala Gly Gly Ala Thr Ala
204      755      760
206 <210> SEQ ID NO: 3
207 <211> LENGTH: 1906
208 <212> TYPE: DNA
209 <213> ORGANISM: Artificial
211 <220> FEATURE:
212 <223> OTHER INFORMATION: cDNA
214 <400> SEQUENCE: 3
215 agaatggagc cctcctggct tcaggaactc atggctcacc ccttcttgct gctgatactc 60
217 ctctgcatgt ctctgctgct gtttcaggta atcaggttgt accagaggag gagatggatg 120
219 atcagagccc tgcacctggt tctgcacccc cctgcccact ggttctatgg ccacaaggag 180

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/540,310

DATE: 06/30/2005

TIME: 08:45:07

Input Set : A:\689290-244.txt

Output Set: N:\CRF4\06302005\J540310.raw

```

221 ttttaccag taaaggaggt tgaggtgtat cataagctga tggaaaaata cccatgtgct 240
223 gttcccttgt gggttggacc ctttacgatg ttcttcagtg tccatgaccc agactatgcc 300
225 aagattctcc tgaaaagaca agatcccaaa agtgctgtta gccacaaaat ccttgaatcc 360
227 tgggttggtc gaggacttgt gaccctggat gggtctaaat ggaaaaagca ccgccagatt 420
229 gtgaaacctg gcttcaacat cagcattctg aaaatattca tcacatgat gtctgagagt 480
231 gttcggatga tgctgaacaa atgggaggaa cacattgccc aaaactcacg tctggagctc 540
233 tttcaacatg tctccctgat gaccctggac agcatcatga agtgtgcctt cagccaccag 600
235 ggcagcatcc agttggacag taccctggac tcatacctga aagcagtgtt caaccttagc 660
237 aaaatctcca accagcgcct gaacaatttt ctacatcaca acgacctggg tttcaaattc 720
239 agctctcaag gccaaatctt ttctaaattt aaccaagaac ttcatcagtt cacagagaaa 780
241 gtaatccagg accggaagga gtctcttaag gataagctaa aacaagatac tactcagaaa 840
243 aggcgtggg attttctgga catacttttg agtgccaaaa gcgaaaacac caaagatttc 900
245 tctgaagcag atctccaggc tgaagtgaac acgttcatgt ttgcaggaca tgacaccaca 960
247 tccagtgtca tctcctggat cctttactgc ttggcaaaagt accctgagca tcagcagaga 1020
249 tgccgagatg aaatcaggga actcctaggg gatgggtcct ctattacctg ggaacacctg 1080
251 agccagatgc cttacaccac gatgtgcac aaggaatgcc tccgcctcta cgcaccggtg 1140
253 gtaaacatat cccggttact cgacaaaccc atcaccttcc cagatggacg ctcccttacct 1200
255 gcaggaataa ctgtgtttat caatatttgg gctcttcacc acaaccctta tttctgggaa 1260
257 gaccctcagg tctttaaccc cttgagattc tccagggaaa attctgaaaa aatacatccc 1320
259 tatgccttca taccattctc agctggatta aggaactgca ttgggcagca ttttgccata 1380
261 attgagtgtg aagtggcagt ggcattaact ctgctccgct tcaagctggc tccagaccac 1440
263 tcaaggcctc cccagcctgt tctcaagtt gtctcaagt ccaagaatgg aatccacgtg 1500
265 tttgcaaaaa aagtttgcta attttaagtc ctttctgata agaattaatg agacaatttt 1560
267 cctaccaaag gaagaacaaa aggataaata taatacaaaa tatatgtata tgggtgtttg 1620
269 acaaattata taacttagga tacttctgac tggttttgac atccattaac agtaatttta 1680
271 atttctttgc tgtatctggg gaaaccacaa aaaacacctg aaaaaactca agctgacttc 1740
273 cactgcgaag ggaaatattg gtttgtgtaa ctagtggtag agtggttttc aagcatagtt 1800
275 tgatcaaaac tccactcagt atctgcatta cttttatctc tgcaaatatc tgcagatag 1860
277 ctttattctc agttatcttt ccccataata aaaataatat gccaca 1906

```

280 &lt;210&gt; SEQ ID NO: 4

281 &lt;211&gt; LENGTH: 505

282 &lt;212&gt; TYPE: PRT

283 &lt;213&gt; ORGANISM: Artificial

285 &lt;220&gt; FEATURE:

286 &lt;223&gt; OTHER INFORMATION: Putative protein derived from cDNA.

288 &lt;400&gt; SEQUENCE: 4

```

290 Met Glu Pro Ser Trp Leu Gln Glu Leu Met Ala His Pro Phe Leu Leu
291 1 5 10 15
293 Leu Ile Leu Leu Cys Met Ser Leu Leu Leu Phe Gln Val Ile Arg Leu
294 20 25 30
296 Tyr Gln Arg Arg Arg Trp Met Ile Arg Ala Leu His Leu Phe Pro Ala
297 35 40 45
299 Pro Pro Ala His Trp Phe Tyr Gly His Lys Glu Phe Tyr Pro Val Lys
300 50 55 60
302 Glu Phe Glu Val Tyr His Lys Leu Met Glu Lys Tyr Pro Cys Ala Val
303 65 70 75 80
305 Pro Leu Trp Val Gly Pro Phe Thr Met Phe Phe Ser Val His Asp Pro
306 85 90 95
308 Asp Tyr Ala Lys Ile Leu Leu Lys Arg Gln Asp Pro Lys Ser Ala Val

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/540,310

DATE: 06/30/2005

TIME: 08:45:07

Input Set : A:\689290-244.txt

Output Set : N:\CRF4\06302005\J540310.raw

```

309          100          105          110
311 Ser His Lys Ile Leu Glu Ser Trp Val Gly Arg Gly Leu Val Thr Leu
312          115          120          125
314 Asp Gly Ser Lys Trp Lys Lys His Arg Gln Ile Val Lys Pro Gly Phe
315          130          135          140
317 Asn Ile Ser Ile Leu Lys Ile Phe Ile Thr Met Met Ser Glu Ser Val
318 145          150          155          160
320 Arg Met Met Leu Asn Lys Trp Glu Glu His Ile Ala Gln Asn Ser Arg
321          165          170          175
323 Leu Glu Leu Phe Gln His Val Ser Leu Met Thr Leu Asp Ser Ile Met
324          180          185          190
326 Lys Cys Ala Phe Ser His Gln Gly Ser Ile Gln Leu Asp Ser Thr Leu
327          195          200          205
329 Asp Ser Tyr Leu Lys Ala Val Phe Asn Leu Ser Lys Ile Ser Asn Gln
330 210          215          220
332 Arg Met Asn Asn Phe Leu His His Asn Asp Leu Val Phe Lys Phe Ser
333 225          230          235          240
335 Ser Gln Gly Gln Ile Phe Ser Lys Phe Asn Gln Glu Leu His Gln Phe
336          245          250          255
338 Thr Glu Lys Val Ile Gln Asp Arg Lys Glu Ser Leu Lys Asp Lys Leu
339          260          265          270
341 Lys Gln Asp Thr Thr Gln Lys Arg Arg Trp Asp Phe Leu Asp Ile Leu
342          275          280          285
344 Leu Ser Ala Lys Ser Glu Asn Thr Lys Asp Phe Ser Glu Ala Asp Leu
345          290          295          300
347 Gln Ala Glu Val Lys Thr Phe Met Phe Ala Gly His Asp Thr Thr Ser
348 305          310          315          320
350 Ser Ala Ile Ser Trp Ile Leu Tyr Cys Leu Ala Lys Tyr Pro Glu His
351          325          330          335
353 Gln Gln Arg Cys Arg Asp Glu Ile Arg Glu Leu Leu Gly Asp Gly Ser
354          340          345          350
356 Ser Ile Thr Trp Glu His Leu Ser Gln Met Pro Tyr Thr Thr Met Cys
357          355          360          365
359 Ile Lys Glu Cys Leu Arg Leu Tyr Ala Pro Val Val Asn Ile Ser Arg
360          370          375          380
362 Leu Leu Asp Lys Pro Ile Thr Phe Pro Asp Gly Arg Ser Leu Pro Ala
363 385          390          395          400
365 Gly Ile Thr Val Phe Ile Asn Ile Trp Ala Leu His His Asn Pro Tyr
366          405          410          415
368 Phe Trp Glu Asp Pro Gln Val Phe Asn Pro Leu Arg Phe Ser Arg Glu
369          420          425          430
371 Asn Ser Glu Lys Ile His Pro Tyr Ala Phe Ile Pro Phe Ser Ala Gly
372          435          440          445
374 Leu Arg Asn Cys Ile Gly Gln His Phe Ala Ile Ile Glu Cys Lys Val
375          450          455          460
377 Ala Val Ala Leu Thr Leu Leu Arg Phe Lys Leu Ala Pro Asp His Ser
378 465          470          475          480
380 Arg Pro Pro Gln Pro Val Arg Gln Val Val Leu Lys Ser Lys Asn Gly
381          485          490          495

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/10/540,310

DATE: 06/30/2005  
TIME: 08:45:08

Input Set : A:\689290-244.txt  
Output Set: N:\CRF4\06302005\J540310.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:1; N Pos. 517,588,597,598

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete, per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:3,4

**VERIFICATION SUMMARY**

DATE: 06/30/2005

PATENT APPLICATION: US/10/540,310

TIME: 08:45:08

Input Set : A:\689290-244.txt

Output Set: N:\CRF4\06302005\J540310.raw

L:10 M:270 C: Current Application Number differs, Replaced Current Application Number  
L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:48 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1 after pos.:480  
M:341 Repeated in SeqNo=1